The test effect: Behavioral change and potential biases due to (biomedical) testing in surveys[☆]

July 25, 2017

Zlata Tanović[a,b,c,*]

[a] *Vrije Universiteit Amsterdam, Faculty of Economics and Business Administration, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands*
[b] *Universitat Pompeu Fabra, Department of Economics and Business, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain*
[c] *Amsterdam Institute for International Development, Pietersbergweg 17, 1105 BM Amsterdam, The Netherlands*

Abstract

Does biomedical testing change health care seeking behavior of the research sample, and can it bias impact estimates of a related health care intervention? This paper is the first to rigorously analyze unintended effects of biomedical testing in surveys. Random assignment of blood pressure measurements in a 2013 household survey in Tanzania, and a second survey of the same individuals two years later, allows for the identification of this "test effect" on health care provider consultations for hypertension (chronic high blood pressure) and uptake of voluntary health insurance. As these were the baseline and follow-up surveys of a health insurance impact evaluation, the possible bias in the insurance impact estimates caused by the test effect can also be estimated. Since, complying with ethical standards, respondents who were tested were told their test result, the differential effect of high versus normal measured blood pressure can be determined. Having high measured blood pressure significantly increased the likelihood of health care provider consultations for hypertension. No evidence is found of a test effect for health insurance enrollment, nor of a bias in health insurance impact estimates due to the blood pressure measurements, for any of the outcomes.

Keywords: test effect, survey methodology, biomedical test, blood pressure, health

[☆]I am grateful to the attendees of the 2016 CSAE conference, the 2015 NEUDC conference, Joachim de Weerdt, Respichius Mitti, Zoë McLaren, Tobias Rinke de Wit, Daniëlla Brals, Marleen Hendriks, Heleen Nelissen, Constance Schultsz, Chris Elbers, Wendy Janssens, Menno Pradhan, Melinda Vigh, Jacques van der Gaag, Martin Wiegand, for useful comments. Many thanks to Economic Development Initiatives Ltd. and the Amsterdam Institute for Global Health and Development for excellent collaboration in survey design and data collection. I would like to thank the Kilimanjaro Native Co-operative Union, MicroEnsure, PharmAccess Foundation, and the Health Insurance Fund for ample assistance with fieldwork preparations. The Health Insurance Fund funded the data collection and has made the data available for this research, for which I am grateful. I kindly thank the Amsterdam Institute for International Development for funding my time in writing this manuscript. This paper is funded under the grant Policy Design and Evaluation Research in Developing Countries" Initial Training Network (PODER), which is funded under the Marie Curie Actions of the EU's Seventh Framework Programme (Contract Number: 608109).

[*]Postal and visiting address is at Universitat Pompeu Fabra. Tel: +34 93542 1512.
Email address: z.tanovic@vu.nl (Zlata Tanović)

insurance, Tanzania.
JEL: C83, C93, I12, I13, O15.

1. Introduction

As developing countries generally have no population-wide administrative health care data, health economics research requires population representative data to be collected through surveys. These often include self-reported morbidity indicators as measures of health status. However, these subjective health measures suffer from misreporting due to respondents' lack of knowledge of their true health status (Sen, 2002), respondents' tendency to report a socially desirable health status (Latkin and Vlahov, 1998; Adams et al., 2005), or because of recall bias (Das et al., 2012). Instead, researchers can opt to collect objective health measures by including biomedical tests in the survey, such as anthropometric measurements, blood pressure measurements, and blood or saliva tests. While this allows more precise knowledge of a respondent's true health status, biomedical testing in surveys has its own drawbacks. The test may give false positives or false negatives (Banoo et al., 2008), there may be non-response bias due to test refusal (e.g. Reniers and Eaton, 2009; Janssens et al., 2014), and test results can even be faked by interviewers (Janssens et al., 2010).

The current paper is about another potential drawback of biomedical testing in surveys, namely its ability to change a respondent's future health care seeking behavior. Such change in behavior could effectively cause a carefully chosen representative sample to cease being representative after the survey. In a panel setting it may bias research outcomes—such as impact estimates of a health care intervention.[1]

It is well known that being surveyed can change behavior. Knowledge of being observed in the scope of an impact evaluation can cause behavioral change in both the treatment and control group, called the Hawthorne and John Henry effect, respectively (Duflo et al., 2007). The question-behavior effect, which includes self-prophecy and mere-measurement effects, occurs when people are asked questions about future behavior and consequently change their actions in line with their answers (Sherman, 1980; Feldman and Lynch, 1988; Sprott et al., 2006; Dholakia, 2010). The survey effect occurs when asking questions about a certain subject causes a reminder or salience shock that makes people act more responsibly afterward (Bridge et al., 1977; Zwane et al., 2011; Crossley et al., 2014; Axinn et al., 2015).

This paper concerns itself with unintended behavioral changes due to biomedical testing

[1] Note that there may be a positive side to this behavioral change as well. If receiving biomedical measurements causes a rise in health insurance demand among the poor, which has proven notoriously difficult to achieve in developing countries (Gwatkin et al., 2004; Victora et al., 2004; Gwatkin and Ergo, 2011), large-scale biomedical testing may get us one step closer to universal health coverage.

in a survey setting. Disclosing previously unknown information about one's health status closes an information gap, raising awareness of one's true health status, rather than being a reminder of something already known. This effect, which I will call "test effect", is thus intrinsically different from those previously mentioned. It is expected to occur when health implications of a revealed test result can be mitigated by behavioral change, such as health care use or healthy living.[2] Non-biomedical tests can fall into this category as well if revealing the test result closes an information gap on a personal health related issue. For example the quality of a household's drinking water directly affects the health of its members, and receiving information that the water is contaminated may lead to better hygiene and water management practices. *Rapid* biomedical tests—where the interviewer can see the test result—are special, because the interviewer is obliged to report the test result to the respondent for ethical reasons. If notification of the test result changes the respondent's behavior, and subsequently sample representativeness, such rapid biomedical measurements can have unintended (and unwanted) consequences for research.

One important concern rises in the context of an impact evaluation of a related health care intervention, when biomedical data are collected during a baseline survey, after which one part of the research population gains access to the intervention, while the other part remains untreated. For example—as is the case in this paper—when evaluating a health insurance program and collecting baseline data on blood pressure (BP), notifying someone that their BP is high might increase that individual's likelihood to take up health insurance,[3] enabling frequent health care utilization, and ultimately improvement in health outcomes. In the control group however, for lack of health insurance, individuals who learn they have high BP might not be able to afford the health care they need. Conversely, learning one has normal BP may decrease their doctor visits and probability to take up health insurance. If access to the insurance indeed makes the intervention group behave differently from the control group, the test effect would bias the insurance impact estimates on health care utilization and health outcomes. When conducting such an impact evaluation it is not common practice to randomize who receives the biomedical measurement at baseline, making it difficult for the researcher to disentangle this bias from the true impact estimate.

In the experiment of this paper 80% of sampled households were randomly assigned to receive BP measurements at baseline, in the scope of an impact evaluation of a voluntary health insurance program in the Kilimanjaro region of Tanzania. Half of the sample gained access to the health insurance program—which included hypertension (HT) treatment—five

[2]Height measurement for example is not expected to change behavior.

[3]Determining whether someone has *hypertension*, i.e. chronically high BP, is possible after high BP is measured multiple times over the span of several days. During the baseline survey individual BP was measured on the same day, which is insufficient for diagnostic purposes.

to nine months after the baseline survey. With these data it is possible to assess the effect of the BP measurements on health care seeking behavior, including take-up of health insurance, in a panel household survey setting.[4] The effect of a normal and high BP result on these outcomes can be differentiated, and the bias in the insurance intervention impact estimates for these outcomes can be estimated. To the author's knowledge, this paper is the first to rigorously analyze unintended effects of biomedical testing in surveys.

Evaluating a health insurance intervention in rural Nigeria, Hendriks et al. (2014) seem to detect a test effect with BP measurement. The authors note that individuals who had high BP at baseline were more likely to use anti-hypertensive drugs, and had lower average BP two years later. This was observed both in the insurance intervention and in the control area—where the health insurance was not available. However, for lack of a test control group, the authors could not assess whether the baseline measurements caused this behavioral change or whether they biased the health insurance impact estimates. Tarozzi et al. (2014) encounter an unintended test effect in rural India, where a rapid diagnostic malaria test was included in the baseline survey. Receiving an offer of insecticide treated bednets approximately four months later, demand was observed to be higher in households where at least one member had tested positive for malaria at baseline.

Several authors have rigorously analyzed how behavior changes when individuals learn their HIV status (Thornton, 2008, 2012; Delavande and Kohler, 2012; Gong, 2015) or their household's water quality (Jalan and Somanathan, 2008; Davis et al., 2011; Luoto et al., 2011; Hamoudi et al., 2012) in the scope of a survey. These authors generally find heterogeneous behavioral effects of—using the terminology of Jalan and Somanathan (2008)—receiving "good" versus "bad" news.[5] Furthermore, Gong (2015) observes that the behavioral response occurs when the test result reveals unexpected information, i.e. that which is contrary to prior beliefs. However these experiments did not use rapid tests—thus notification of the test result was optional in principle—and none look at how the behavioral change due to testing affects impact estimates of a related intervention.[6] Evidence is lacking on these unintentional test effects in particular.

The paper will proceed as follows. The next section provides information on the research population and the health insurance intervention. The experiment and data collection are described in Section 3. Descriptive statistics are given in Section 4. In Section 5 the model is presented, followed by the econometric analysis in Section 6. Section 7 concludes.

[4] Bias in measured BP, the health outcome of interest, cannot be assessed however, because—due to time and budget constraints—BP at follow-up was measured only for those with high BP at baseline.

[5] See Appendix A for a short summary of these papers.

[6] Zwane et al. (2011) did explore such unintended *survey* effects. They found that frequent surveying biased the impact estimate of protected water sources on child diarrhea. They however detected no bias in health insurance price sensitivity due to surveying.

2. Health insurance intervention

The background of the surveys is an impact evaluation of the KNCU Health Plan, a voluntary subsidized health insurance program in a rural part of the Kilimanjaro region of Tanzania—on the slopes of mount Kilimanjaro. It was funded by the Health Insurance Fund,[7] and implemented by the PharmAccess Foundation (Dutch NGO) and MicroEnsure (an insurance company operating i.a. in Tanzania). This health insurance was offered to small scale coffee farmers who are active members of the Kilimanjaro Native Cooperative Union (KNCU), and their household members.[8] Insured individuals could go with their insurance card to a designated primary health facility in their vicinity—usually an upgraded faith based dispensary—where they would receive free outpatient treatment. Referrals to a district hospital were covered for pregnancy related complications only. The insurance was available at the household—rather than individual—level, and was on an annual basis.[9] The annual co-premium had to be paid in advance in cash and was priced on a sliding scale, ranging from 12,000 Tanzanian Shilling (TZS) for one person, to 45,000 TZS for 9–12 people per household.[10]

KNCU is organized into about 90 primary societies (PSs), which are spread throughout the Kilimanjaro region (KNCU). Because the KNCU Health Plan was gradually expanded by PS since April 2011, the insurance intervention and control groups consisted of several PSs in rural districts Hai, Moshi Rural, and Rombo: five PSs in the insurance intervention group, and four in the control group. These were chosen such that they were far enough apart to prevent spill-overs from the intervention to the control group, and such that they were similar at baseline on key characteristics such as access to health facilities, altitude, distance to Moshi town, type of coffee grown, and in which district they were located. A map of the insurance intervention and control area is shown in Figure C1 of the Appendix.[11]

Five to nine months after the baseline survey, between July and October 2013, the KNCU Health Plan was offered to the insurance intervention group in districts Hai and Moshi Rural only. It proved not possible to introduce the insurance in Rombo district, which is why

[7] See www.hifund.org for more information.

[8] Based on interviews with local village leaders, approximately two out of five households in the study area are estimated to be such "KNCU households", ranging from 10% to 89%, depending on the village.

[9] Note however that it was difficult for the insurer to check how many individuals resided in the household at the time of insurance, because people would claim that certain household members had moved out since the census (conducted in November 2012 to January 2013, see Appendix B). In households that enrolled in the KNCU Health Plan, on average 84% of all household members who did not move out since the baseline survey were covered by the insurance at first enrollment. 63% of these households were fully insured, i.e. all household members who did not move away since baseline were insured. There is no clear negative relationship between household size and the percentage of insured individuals per household: the correlation coefficient, after removing one exceptionally large household with 11 individuals, is -0.16 (p=0.170).

[10] Since the full premium was 14,000 TZS per person, this implies a subsidy level of 14% up to 73%, depending on household size.

[11] For more detailed information on the insurance intervention and control group choice and similarities at baseline, see AIID and AIGHD (2013).

subsequently the insurance intervention and control primary society in this district were excluded from the KNCU Health Plan impact evaluation—and from the follow-up survey.[12]

There is heterogeneity in the KNCU Health Plan roll-out between the Hai and Moshi Rural PSs (see Table C1 of the Appendix). The two insurance intervention PSs in Hai (Moshi Rural) district received access to the insurance in June–July (October) 2013, and had an upgraded (non-upgraded) dispensary as primary KNCU Health Plan facility. Furthermore, in Hai district the KNCU Health Plan could be renewed after 12 months, while the two Moshi Rural primary societies received free extension until the end of 2014, after which re-enrollment was not possible. Namely, in that district, the KNCU Health Plan was replaced by a new insurance since the start of 2015, as will be described in this section's final paragraph. Heterogeneous test effects between the two districts will be explored in Section 6.

The study population in the districts Hai and Moshi Rural is quite poor, with median daily per capita consumption at 2000 TZS or 1.25 USD (3.13 USD at PPP) at baseline.[13] The full KNCU Health Plan annual per capita premium thus amounts to one week of median per capita consumption. At baseline, 10.8% of the survey sample had health insurance; 8.5% were insured by the National Health Insurance Fund (NHIF), and 2.1% by the Community Health Fund (CHF). CHF is a community based health insurance, available to the full district population, and managed by the district government. Its co-premium is 10,000 TZS per household per year, and covers up to six household members. Insured individuals are entitled to outpatient treatment in public health facilities in the district. However, treatment is not always free in practice. NHIF is the government health insurance, mandatory for government employees (NHIF). The co-premium is 3% of income, and is half of the full premium. It covers all health care in public, and selected private, health facilities. The spouse and up to four legal dependents are eligible as beneficiaries. NHIF is available to informal sector employees for 987,500 TZS ($\approx$ 617 USD) per family per year, more than the median annual per capita consumption in the research area (730,600 TZS). The KNCU Health Plan was introduced to provide high quality health care at a low price, hypothesizing that this was not available to the target population—outside of the public sector. Note that HT treatment is included in all three insurance programs.

In Moshi Rural district, the KNCU Health Plan joined with CHF as of 1 January 2015, in a public-private partnership with the local governments of these districts. It was re-named *improved* Community Health Fund (*i*CHF), and is now available to the full district population.

[12]This loss of sub-population is not expected to have caused sample selection, since both the choice of insurance intervention and control group as well as the sampling was stratified by area (Rombo vs. non-Rombo district), see AIID and AIGHD (2013) and Appendix B.

[13]Source: baseline survey. 1 USD $\approx$ 1,600 TZS in February 2013 (Oanda). 1 USD $\approx$ 0.4 USD at Purchasing Power Parity (PPP) in Tanzania in 2013 (World Bank).

Note that the insurance control group is located in Hai district. At the time of the follow-up survey the marketing of *i*CHF had not yet reached the research population—only one household in the treatment group had heard of it, and three in the control group—and none of the surveyed individuals were insured by *i*CHF at the time of the follow-up survey in March 2015. In May of that same year—one month after completion of all data collection—*i*CHF became available to the Hai district population as well.[14]

3. Experimental design and data collection

The baseline survey was conducted between 25 January to 6 March 2013 by Economic Development Initiatives (EDI) Ltd., a Tanzanian survey firm. Six teams totaling 24 interviewers and seven health officers conducted household interviews and biomedical tests, respectively, in 1,500 KNCU households—half of which belonged to the insurance intervention group. The random sample was stratified at the sub-village level.[15] Interviewers were trained to introduce themselves as being from an independent survey firm, working together with two Amsterdam research institutes—the Amsterdam Institute for International Development (AIID) and the Amsterdam Institute for Global Health and Development (AIGHD)—performing research on health insurance. The KNCU Health Plan was not mentioned. To check whether respondents still anticipated that the survey was part of KNCU Health Plan related research, one of the open questions posed was—after explaining the concept of health insurance—which health insurances respondents had heard of. Only 6% of households had heard of the KNCU Health Plan—8% in the insurance intervention area, and 4% in the insurance control area—lessening concerns of an anticipation effect.

The household questionnaire was conducted in Swahili, the lingua franca of Tanzania,[16] using computers with the specialized survey software Surveybe. It was very extensive, containing sections on education, work, consumption, household assets, gifts and loans, coffee production, risk and time preferences, self-reported health status, health care expenditures, and health care seeking behavior. All individuals, including those not assigned to receive BP measurements, were asked to give written consent for the biomedical part of the interview.[17] Complying with ethical standards, without this consent biomedical measurements and biomedical questions were fully excluded from the interview.

Furthermore, survey medical officers from EDI Ltd. were assigned to perform BP measurements on all adults in randomly selected 80% of sampled households—stratified by sub-

[14] Siha was the first Kilimanjaro district where *i*CHF started, in November 2014.
[15] See Appendix B for more sampling details.
[16] Only for six households translation was needed to Chagga, the area's tribal language.
[17] Either the caretaker or the household head was asked to give consent for minors.

village—using the OMRON M6 Comfort digital BP device.[18] These were extensively trained by medical doctors from AIGHD to be able to perform the measurements according to the highest standards. To signal professionalism of the procedures, the survey medical officers were dressed in doctors' white coats.[19]

If an individual had high BP (systolic BP $\geq$ 140 mmHg or diastolic BP $\geq$ 90 mmHg) in two out of three measurements, the medical officer would point out the health risks of high BP with the help of a leaflet especially designed for the survey, and would advise the person to visit a health care professional for additional testing and treatment.[20] The leaflet, written in Swahili, was finally handed to the respondent for their information.

Two years later, in March 2015, EDI Ltd. returned to interview the same households. Because Rombo district was not anymore included in the follow-up survey (see Section 2), the sample was reduced from 1,500 to 1,000 households. The follow-up survey questionnaire was less extensive than that of the baseline, but included questions on self-reported health status, health care utilization, health expenditures, and detailed health insurance questions—including health insurance status one year before the follow-up survey, in March 2014.

Medical ethical clearance for both survey rounds was received from the Tanzania National Institute for Medical Research (NIMR). The Tanzania Commission for Science and Technology (COSTECH) gave general research clearance.

4. Data description

The 1,000 baseline households from districts Hai and Moshi Rural had 4,122 household members in total, out of which 2,530 were adults (BP test: 2,038; No BP test: 492)[21]. Out of all adults 2,159 (85%) consented for the biomedical part of the survey (BP test: 1,738 = 86%; No BP test: 421 = 85%). From all who received the BP measurements—namely all consenting adults, except 43 individuals who did give consent but were not available for the measurements—588 (35%) had high BP in at least two out of three measurements, which

[18] Individuals between 12 and 59 years of age in the same subset of households were furthermore assigned to receive lung function measurements. However, since it turned out extremely difficult for respondents to correctly perform these lung function measurements—one needed to blow three times into a machine with considerable strength—only 34% successfully completed the procedure. Furthermore, only 5% of those individuals (2% of the total) showed signs of obstructive pulmonary disease, and were advised to seek medical care for this condition. This is why lung function measurements will be excluded from the analyses. Subsequent results are robust to exclusion of individuals with signs of obstructive pulmonary disease at baseline. Moreover, it should be noted that all consenting individuals in all surveyed households received anthropometric measurements. Because individuals were not warned of health issues concerning anthropometrics, a behavioral response from these measurements is not expected.

[19] On the one hand the white coat is expected to have increased the credibility of the medical officers' health advice. On the other hand, it can cause elevated stress, temporarily increasing BP in the person receiving the measurement, causing "white coat" HT in some individuals (Pickering et al., 1988).

[20] High measured BP is a warning sign of HT, but is not enough to diagnose it (footnote 3).

[21] "BP test" are the individuals in households that were randomly *assigned* to receive BP measurements, and "No BP test" are those in households that were not assigned to receive the measurements.

will just be called "high BP" in the remainder of this paper. If the individual had at most one high BP measurement out of the three, this person is said to have "normal BP"—such individuals were not warned that they are at risk of cardiovascular disease.

Out of these 1,000 baseline households 34 could not be reached at follow-up (BP test: 29 = 4%; No BP test: 5 = 3%), out of which 19 (56%) were in the insurance intervention group. These households had either moved, were unavailable of had refused an interview. Out of the baseline consenting adults 1,800 (83%) were still household members at follow-up (BP test: 1,444 = 83%; No BP test: 356 = 85%) and 1,536 (85%) consented at follow-up (BP test: 1,243 = 86%; No BP test: 293 = 82%).[22] These 1,536 individuals are used in the analyses.

Individuals who left the household or did not give consent at follow-up were on average 12 years younger, more likely to be male, healthier, better educated, less likely to be married, more likely to have worked in the past year, and more likely to have experienced a financial health shock at baseline, compared to those in the research sample (Table C2). The baseline characteristics of those not available for the (biomedical part of the) follow-up survey were however balanced between the BP test and control group (Table C3).

Table 1 compares, for all adults who consented for the biomedical part in both surveys, the baseline means of the outcome variables and other control variables, by BP test assignment. Out of the 22 baseline characteristics, three are not balanced at the 5% level, all age related. Individuals not assigned to receive BP measurements are found to be on average three years older than those assigned to receive the measurements.[23] Since HT risk increases with age, this must be taken into account in the analyses. Consequently observations will be reweighted according to age in the robustness checks.

Baseline characteristics are balanced between the insurance intervention and control area, except for high BP prevalence, which is almost one and a half times higher in the insurance control area than in the insurance intervention area (Table C4).[24] Since observed test results can be controlled for, this imbalance is not problematic for analyses.

Figure 1 shows the means of the outcomes of interest over time, disaggregated by test result and by the insurance intervention and control area. Next to health insurance enrollment, two health care utilization outcomes are considered—whether one had a BP check in the past 12 months,[25] and whether the person consulted a health care provider for HT, over the same

[22] Only 25 of those had a missing test result at baseline.

[23] Looking at the mean age separately for the insurance intervention and control group it turns out that the imbalance is small and not statistically significant in the intervention area (BP test: 54.7 years; No BP test: 55.7 years; p=0.479). However the difference in means is quite severe in the insurance control area (BP test: 54.9 years; No BP test: 60.1 years; p=0.008). When regressing BP test assignment on age-group dummies, insurance intervention area, and their interactions, it becomes clear that the oldest group (70 years and older) is overrepresented in the *non-tested* individuals of the *insurance control* area.

[24] This is not due to age imbalance, since regressing the insurance intervention area on age-group dummies for those who received BP measurements yields a p-value of 0.499 for the joint significance F-test.

[25] Individuals were shown pictures of several BP measurement devices upon asking this question.

Table 1: Means of baseline variables, by BP test assignment

	BP test Mean (N=1243)	No BP test Mean (N=293)	ΔMean p-value
Main			
Insurance intervention area	0.48	0.54	0.186
Self-reported HT	0.23	0.26	0.189
BP check - past 12 months	0.34	0.37	0.467
Consult for HT - past 12 months	0.16	0.19	0.229
Any health insurance	0.15	0.13	0.367
Socio-economic characteristics			
Age (years)	54.8	57.7	0.016*
Female	0.61	0.59	0.476
Married[a]	0.69	0.70	0.686
Worked[b] - past 12 months	0.21	0.17	0.073^{+}
Religion: Christian	0.96	0.96	0.887
Mother tongue: Chagga	0.99	0.97	0.250
Educ[c]: None	0.09	0.13	0.094^{+}
Educ: Less than primary school	0.31	0.32	0.827
Educ: Primary school	0.54	0.49	0.163
Educ: More than primary school	0.06	0.06	0.853
Self-reported illness/ injury			
Chronic illness	0.41	0.46	0.157
Acute illness / injury - past 12 months	0.50	0.52	0.553
Hospitalization - past 12 months	0.07	0.08	0.546
Household characteristics			
Annual consumption[d] - PC (TZS/1,000)	860	872	0.742
Financial health shock - past 12 months	0.37	0.39	0.632
Household size	4.40	4.09	0.073^{+}
#Young children in HH (age < 5)	0.20	0.12	0.005**
#Elderly in HH (age ≥ 60)	0.70	0.78	0.199
#Reproductive age women (15–45) in HH	0.51	0.39	0.042*

Note: The table shows statistics for all adults who gave consent for the biomedical part of both surveys (questions and measurements). Means are weighted and p-values clustered at the household level, in accordance with the sampling method. HH=household; PC=per capita; [a]Includes mono- and polygamous marriage; [b]Apart from household chores and family farming; [c]Highest completed education level; [d]One outlier excluded; $^{+}$p<.10, *p<.05, **p<.01, ***p<.001.

time period.[26] Additionally, self-reported HT is considered as well, because subjective beliefs of one's HT status are a key part in the mechanism that leads from notification of the BP test result to changing one's health care seeking behavior.

The general picture from Figure 1 is that—as expected—already at baseline, individuals with high BP were more likely to self-report HT, more likely to have utilized health care for HT in the past year, and more likely to have health insurance, compared to those with normal BP at baseline. Baseline means of individuals from the test control group—residing in households not assigned to receive BP measurements—are a priori expected to be in between those with normal and high BP. This can indeed be seen in the insurance intervention group, for all outcomes except for the insurance prevalence, which is lower than expected for those in the test control group. However, in the insurance control group, the baseline means are as high or higher than those of individuals with high BP at baseline. This is an indication that HT prevalence in the test control group may very well be higher than that of the population,

[26]Note that the follow-up survey occurred 25–26 months after the baseline survey. Thus if during the follow-up survey someone reports to have utilized health care in the past 12 months, this visit occurred at least one year after the baseline survey.

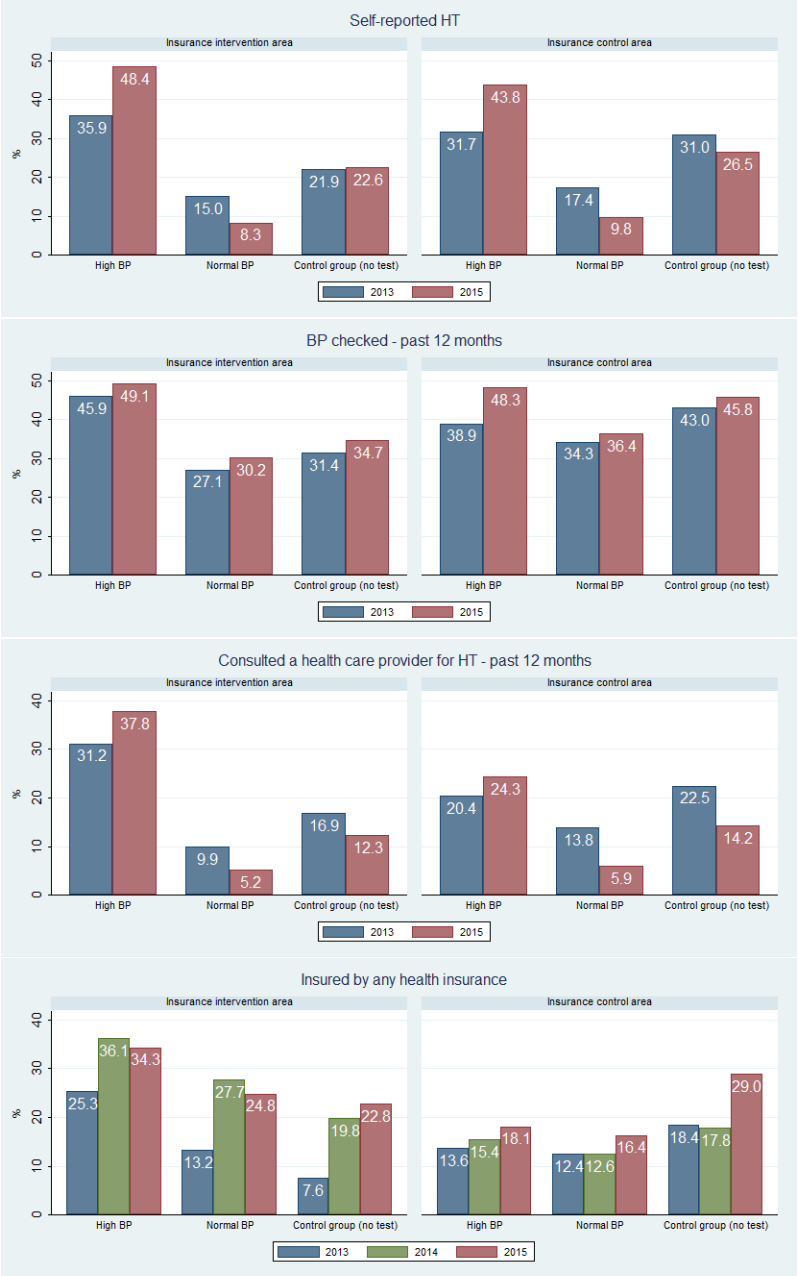

Figure 1: Outcomes by test result and insurance area (N=1511). Observations in the insurance intervention [control] area: Total: 789 [722]; High BP: 209 [258]; Normal BP: 425 [326]; Control group (no test): 155 [138].

which is in line with the earlier observation that non-tested individuals in the control group are on average older than those assigned to be tested (footnote 23). Reweighting according to age should control for this imbalance.

From the first set of graphs in Figure 1, it can be seen that after two years, as expected, the percentage of self-reported HT increased (decreased) in both the insurance intervention and control area among those who had high (normal) BP at baseline. Among those who were not assigned to receive BP measurements the percentage of self-reported HT surprisingly

decreased in the insurance control area (statistically insignificant at a 10% level). As can be seen from the second set of graphs, there is a small increase between baseline and follow-up in the percentage of individuals who had their BP checked in the past 12 months, for all subgroups.[27] The third set of graphs, showing the percentage of individuals who consulted a health care provider for HT in the past 12 months, displays a pattern similar to self-reported HT: an increase over time for those with high BP at baseline, and a decrease for those with normal BP. For the test control group there is an unexpected decrease in likelihood of consultations for HT over time, for both the insurance intervention and control group.

The final set of graphs in Figure 1 shows the percentage of individuals enrolled in any health insurance program. As mentioned in Section 2, already at baseline almost 11% of the surveyed individuals had health insurance (2% CHF, 9% NHIF). In the insurance intervention area there is a rise in health insurance in 2014 and 2015, irrespective of the BP measurement. In the insurance control group there is hardly any change in health insurance enrollment between the 2013 baseline survey and 2014, but there is a rise in 2015, for all three groups. The rise in insurance take up in the insurance intervention group is mainly due to uptake of the KNCU Health Plan,[28] while in the control group it is CHF insurance take up that increased between baseline and follow-up (Figure C2).[29] There is no indication that those with high BP are more likely to take up health insurance than those with normal BP. This could be because the KNCU Health Plan is a household level insurance: healthy household members become insured along with the sick.[30]

Finally, the age imbalance between the test and no test group is problematic if the test effect is heterogeneous by age. However, there is no indication of this (Figure C3).

5. Model

For the kth outcome of interest, Y_k, the test effect—irrespective of its result—is captured by parameter β_k in the following difference-in-differences individual fixed effects model:[31]

[27] The household survey BP test was explicitly excluded when phrasing this survey question.

[28] Note that the KNCU Health Plan was introduced there in July–October 2013. Because one KNCU Health Plan insurance package lasts for one year—except in the Moshi Rural insurance intervention area, where it was extended to 15 months (Table C1)—this means that the (March) 2014 insurance status reflects the first-time KNCU Health Plan insurance choice. The (March) 2015 insurance status does *not* reflect the KNCU Health Plan second-time insurance choice well, because there was no second-time enrollment possible in Moshi Rural due to the introduction of the new *i*CHF insurance (Section 2).

[29] NHIF insurance shows no substantial changes over time, which is to be expected for a government employee insurance that is extremely expensive for the informal sector (Section 2).

[30] CHF is a household level insurance as well, while NHIF is not per se (Section 2).

[31] Equation 1 is a linear probability model, since all outcomes of interest are binary. A non-linear model is in theory preferred, e.g. the "changes-in-changes" model by Athey and Imbens (2006), as suggested by Blundell and Dias (2009). However, as shown by Angrist and Pischke (2009, p. 197–205) and Wooldridge (2002, p. 472), estimated marginal effects and standard errors from a non-linear model are generally similar to those of its linear counterpart.

$$y_{kit} = \alpha_k + \beta_k(M_i \times T_t) + \gamma_k T_t + \delta_{ki} + \epsilon_{kit}, \quad (1)$$

where y_{kit} is the kth outcome value for individual i at time t. Time t is equal to 0 at baseline (January–February 2013), and 1 at follow-up (March 2015) except for health insurance uptake. When considering insurance uptake t will be equal to 1 in March *2014* (one year before the follow-up survey), because this reflects the *first-time* KNCU Health Plan choice.[32] Parameter α_k is the constant term and M_i is a dummy equal to 1 if individual i was assigned to receive BP measurements at baseline,[33] and 0 otherwise. T_t is the time dummy—equal to 0 at baseline and 1 at $t = 1$. Parameter β_k thus captures the intent to treat (ITT) of the BP measurement. Because all except 25 of these individuals received the measurements upon consenting, β_k captures the average treatment effect on the treated (ATET) as well. Parameter γ_k gives the common time trend, individual time invariant characteristics are captured in the fixed effect δ_{ki}, and ϵ_{kit} is the error term.[34]

Because the BP measurement is combined with notification of the result—as is necessary to obtain ethical clearance for rapid biomedical testing in surveys—it is expected that the test *result* will drive behavior, rather than the act of testing.[35] To differentiate between the effects of normal and high BP M_i is split up in $M_i = N_i + H_i$, where N_i and H_i are dummy variables for normal and high BP, respectively.[36] The parameters of interest are then β_{kn} and β_{kh} in the following equation:[37]

$$y_{kit} = \alpha'_k + \beta_{kn}(N_i \times T_t) + \beta_{kh}(H_i \times T_t) + \gamma'_k T_t + \delta'_{ki} + \epsilon'_{kit}. \quad (2)$$

The test effect is heterogeneous by test result iff $\beta_{kn} \neq \beta_{kh}$. A priori it is expected that $\beta_{kh} \geq 0$ for all outcomes: individuals hearing that they have high BP will become more

[32] See footnote 28. The insurance status in March 2015 unfortunately does not reflect the second-time KNCU Health Plan choice well, which is why it will not be used in the analyses.

[33] This is thus excluding individuals who did not give consent for the biomedical part of the survey, and including the 25 individuals who gave consent, but did not receive the test.

[34] Estimating the constant term α_k and the individual fixed effects δ_{ki} requires a normalization (there is one degree of overidentification), such as $\sum_i \delta_{ki} = 0$, which is standard for the fixed effects estimator "xtreg, fe" in Stata.

[35] Since there is no intermediate control group which receives the test, but does not learn its outcome, the effect of administering the test only—without letting the respondent know its outcome—cannot be determined.

[36] This split is impossible for the 25 individuals in the treatment group who gave consent but were not tested, thus their N_i and H_i value is set to missing.

[37] Controlling for high BP in both the test treatment and control group—just as Jalan and Somanathan (2008) control for "bad news"—is not possible due to lack of a control group that does not learn the test result (footnote 35). Because the testing treatment was randomly assigned, the a priori expectation is that high BP is balanced between the test treatment and control group. From the descriptive results in Table 1 it seems however that the high BP prevalence in the test control group is likely higher than that of the treatment group, since the latter individuals are on average younger. Fixed effects regression should control for this imbalance as long as the treatment effect is not heterogeneous by age. As mentioned at the end of Section 4, there is no indication of heterogeneity by age in the outcomes of interest. Any remaining age imbalance will be corrected for by reweighting regressions by age in the robustness analyses.

aware of their possible HT status, are more likely to seek care for HT, and are more likely to take up health insurance afterward. Symmetrically, it is expected that $\beta_{kn} \leq 0$ for all outcomes, except health insurance uptake, which could still be attractive, since it protects against many types of health risks—not only HT. Moreover, in case that health insurance is at the household level, as with the KNCU Health Plan and CHF, healthy household members are expected to take up health insurance along with the sick.

Adding to equation 1 interaction terms with dummy D_i, denoting whether individual i resides in the *insurance* intervention area, gives:

$$\begin{aligned} y_{kit} = \tilde{\alpha}_k + \tilde{\beta}_k(M_i \times T_t) + \tilde{\eta}_k(M_i \times D_i \times T_t) + \\ \tilde{\theta}_k(D_i \times T_t) + \tilde{\gamma}_k T_t + \tilde{\delta}_{ki} + \tilde{\epsilon}_{kit}. \end{aligned} \tag{3}$$

Parameter $\tilde{\theta}_k$ captures the insurance intervention ITT on the kth outcome, expected to be non-negative for all outcomes,[38] and distinctly positive for insurance uptake. In the case of the insurance impact, the ATET is not captured by the same parameter, because take-up of the KNCU Health Plan is voluntary and not highly prevalent. The main parameter of interest in equation 3 is $\tilde{\eta}_k$, since it represents the bias from the BP test (disregarding its result) in the insurance intervention ITT estimate.[39] Note finally that $M_i \perp D_i$, because the randomized assignment of the test treatment (M_i) was equally divided between the insurance intervention and control area by design.

The bias may be heterogeneous by test result. Writing $M_i = N_i + H_i$ in equation 3 gives:

$$\begin{aligned} y_{kit} = \tilde{\alpha}'_k + \tilde{\beta}_{kn}(N_i \times T_t) + \tilde{\beta}_{kh}(H_i \times T_t) + \tilde{\eta}_{kn}(N_i \times D_i \times T_t) + \\ \tilde{\eta}_{kh}(H_i \times D_i \times T_t) + \tilde{\theta}'_k(D_i \times T_t) + \tilde{\gamma}'_k T_t + \tilde{\delta}'_{ki} + \tilde{\epsilon}'_{kit}, \end{aligned} \tag{4}$$

where the parameters of interest are $\tilde{\eta}_{kn}$ and $\tilde{\eta}_{kh}$, denoting the bias in the insurance intervention ITT estimate due to a normal and high BP result, respectively. For individuals

[38] An increase in self-reported HT in the insurance intervention area would be in line with recent literature showing evidence that more health care utilization can lead to worse self-reported health status, due to increased knowledge of one's true health status (Dow et al., 2003).

[39] This can be seen as follows. Defining $\Delta y_{ki} := y_{ki1} - y_{ki0}$, omitting the subscript k, and writing M_i (D_i) instead of $M_i \times T_t$ ($D_i \times T_t$) for ease of notation, we have:

$$\begin{aligned} E(\Delta y_i | M_i, D_i) &= C + E(\Delta \epsilon_i | M_i, D_i), \\ &= C + M_i D_i E(\Delta \epsilon_i | M_i = 1, D_i = 1) + (1 - M_i) D_i E(\Delta \epsilon_i | M_i = 0, D_i = 1) \\ &+ M_i(1 - D_i) E(\Delta \epsilon_i | M_i = 1, D_i = 0) + (1 - M_i)(1 - D_i) E(\Delta \epsilon_i | M_i = 0, D_i = 0), \\ &=: C + M_i D_i E_{11} + (1 - M_i) D_i E_{01} + M_i(1 - D_i) E_{10} + (1 - M_i)(1 - D_i) E_{00}, \\ &= C + E_{00} + M_i(E_{10} - E_{00}) + D_i(E_{01} - E_{00}) + M_i D_i[(E_{11} - E_{10}) - (E_{01} - E_{00})], \end{aligned}$$

where C is a constant. The term $(E_{01} - E_{00})$ is the ITT of the insurance program in absence of BP measurements; $(E_{10} - E_{00})$ is the ITT of the BP measurements without the availability of the insurance program; and $[(E_{11} - E_{10}) - (E_{01} - E_{00})]$ is the ITT of the insurance program in the presence of BP measurements, minus the ITT in absence of the BP measurements, i.e. the bias in the insurance program ITT due to the BP measurements. Note finally that in a two period panel the within estimator is equivalent to the first difference estimator up to a constant.

with high measured BP at baseline, access to the KNCU Health Plan is a priori expected to further amplify health care use and insurance uptake, thus $\tilde{\eta}_{kh}$ is expected to be non-negative for the two health care utilization outcomes, as well as for insurance uptake. However, $\tilde{\eta}_{kh}$ may be (close to) zero because (i) household level insurance reduces selection into insurance, (ii) insured individuals do not increase health care use,[40] or (iii) when the KNCU Health Plan became available in the insurance intervention area—five to nine months after the baseline survey—results of the BP measurements were not prominent in people's minds. In case of self-reported HT the sign of $\tilde{\eta}_{kh}$ is unclear, since the baseline high BP may have been acute instead of chronic (see footnote 3). No additional effect of the insurance intervention is expected in case of normal BP, thus $\tilde{\eta}_{kn}$ is expected to be zero.

Finally, the behavioral response to the test is expected to differ conditional on prior beliefs of HT status, i.e. self-reported HT at baseline. In particular, as in Gong (2015), it is expected that the behavioral response will be strongest in those "surprised" with normal/high BP, than in those who are "unsurprised". These heterogeneous effects will be explored in the next section as well.

6. Analysis

Because of the random assignment of households into receiving the BP tests at baseline a priori it could reasonably be assumed that the time trend of the outcome of interest in absence of the baseline test would be the same between the test and no-test group. Then, using a two period balanced panel, the above equations could be consistently estimated with the fixed effects estimator (Cameron and Trivedi, 2009).

However, from Section 4 is was clear that the randomization was unlucky in terms of age—a risk factor for HT—since the test group is on average younger than the no-test group. The individual fixed effects estimator is still consistent if the test effect is not heterogeneous by age—of which there is no indication, as noted at the end of Section 4. In case there is some age heterogeneity, reweighting observations by baseline age-group would solve the issue. This will be done as robustness check.

All subsequent regressions are performed using the command "xtreg, fe" in Stata 11.2 software (StataCorp, 2009).[41] According to the sampling frame, observations are weighted by their sampling probabilities, and errors are clustered at the household level. Only adults who gave consent for the biomedical part of both surveys are included in the regressions.

[40] Reasons (i) and (ii) could be framed in terms of adverse selection and moral hazard, respectively—negative concepts from the point of view of classical health economics. However, because the goal of the KNCU Health Plan is to enable affordable high quality health care to the those who are otherwise unable to afford it, insuring sick underprivileged individuals and having them make more use of health care may be viewed positively.

[41] This estimator uses the normalization that the sum of the fixed effects is zero.

6.1. Main results

Since the test effect is expected to be heterogeneous by test outcome, potentially with reversed signs for three out of four outcomes, it seems prudent to first look at the results corresponding to equation 2 and 4.

Table 2: Main results, corresponding to equation 2

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
Normal BP	-0.053	-0.003	0.001	0.016
	(0.032)	(0.042)	(0.031)	(0.034)
High BP	0.140***	0.039	0.113**	-0.009
	(0.039)	(0.045)	(0.038)	(0.035)
Constant	-0.018	0.029	-0.064*	0.063*
	(0.028)	(0.034)	(0.027)	(0.028)
Observations	3014	3008	3006	3022
$P(\beta_{kn} = \beta_{kh})$	<.001***	0.260	<.001***	0.289

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted); 12m=12 months; $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 2 shows estimation results corresponding to equation 2. There is evidence of a test effect for health provider consultations for HT in the past 12 months (column 3), through changes in self-reported HT (column 1). There is a 14 pp increase (significant at the 0.1% level) in self-reported HT for those who had high BP at baseline, compared to those who were not tested. There is also a 5 pp decrease in self-reported HT for those with normal BP, though this is not significantly different from zero at the 10% level. Having had high BP at baseline increases the probability to consult a health care provider for HT significantly (at the 1% level), by 11 pp. Normal BP is found to have no influence on consultations for HT, and the near zero point estimate implies that the normal BP group behaves no differently than the group that did not receive BP measurements. Heterogeneous test effects between normal and high BP for these two outcomes are confirmed by the small p-value corresponding to the Wald test of $\beta_{kn} = \beta_{kh}$. Comparing column 2 and 3 of the table, the coefficients for having had BP checks in the past 12 months and having consulted a health care provider for HT are qualitatively similar: the coefficient for normal BP is close to zero, and the coefficient for high BP has a positive sign. However, the magnitude of the latter in case of BP checks is much smaller, and not statistically significant at the 10% level. One explanation for this is that individuals who did not receive the baseline BP measurement, or those who had normal BP at baseline, received BP checks during doctor visits not necessarily *for* HT.

Finally, there is no indication of a test effect on insurance take up (column 4 of Table 2). The coefficients are close to zero, insignificant at the the 10% level, and even their signs are opposite to those expected. Thus, based on these results, there seems to be no selection into

health insurance due to the BP measurements.[42]

There is no evidence of any test effect when *not* differentiating by test result (equation 1), for any of the outcomes (Table C5). The heterogeneous test effects of normal and high BP thus seem to cancel each other out.

Table 3: Main results, corresponding to equation 4

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
Normal BP	-0.029	-0.010	0.001	0.008
	(0.048)	(0.056)	(0.043)	(0.046)
High BP	0.169**	0.064	0.121*	0.024
	(0.056)	(0.058)	(0.051)	(0.046)
Normal BP × Ins.area	-0.044	0.013	-0.000	0.016
	(0.065)	(0.083)	(0.061)	(0.067)
High BP × Ins.area	-0.053	-0.061	-0.006	-0.037
	(0.079)	(0.089)	(0.077)	(0.070)
Ins.area	0.055	0.001	0.034	0.128*
	(0.056)	(0.067)	(0.054)	(0.055)
Constant	-0.048	0.028	-0.082*	-0.006
	(0.042)	(0.044)	(0.038)	(0.039)
Observations	3014	3008	3006	3022
P($\tilde{\beta}_{kn} = \tilde{\beta}_{kh}$)	<.001***	0.143	0.003**	0.512
P($\tilde{\eta}_{kn} = \tilde{\eta}_{kh}$)	0.891	0.327	0.921	0.281

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; Ins.area= Insurance intervention area; $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 3 shows the estimation results corresponding to equation 4. Even more pronounced than in Table 2, a positive effect can be seen of high BP on self-reported HT (17 pp increase, significant at the 1% level), health care provider consultations for HT (12 pp increase, significant at the 5% level), and having had BP checks in the past 12 months (6 pp increase, though not significant at the 10% level). There is now a positive effect of high BP on health insurance uptake, but it is small (2 pp) and insignificant at the 10% level. The effect of normal BP on the outcomes is consistently close to zero and insignificant at the 10% level. Unsurprisingly, there *is* a 13 pp impact (significant at the 5% level) of offering the KNCU Health Plan on health insurance enrollment in the insurance intervention area.

The insignificant parameter estimates of the interaction terms "Normal BP × Ins. area" ($\tilde{\eta}_{kn}$) and "High BP × Ins. area" ($\tilde{\eta}_{kh}$), as well as the insignificant Wald test of $\tilde{\eta}_{kn} = \tilde{\eta}_{kh}$ for all outcomes, gives no evidence that the test biases the KNCU Health Plan impact estimates. Possibly this is because the KNCU Health Plan is offered at the household level only, thus limiting selection into insurance. If the insurance premium were the same per capita for all, this hypothesis would be consistent with small households with at least one high BP member being more likely to take up insurance than large households. There is however an opposite

[42] When running regression 4 of Table 2 at the household level (replacing the individual high BP with a dummy for high BP in the household, and similarly replacing the individual normal BP with a dummy for only normal BP in the household) the signs and significance of the parameter estimates remain the same.

mechanism as well related to household size, because of the sliding scale pricing of the KNCU Health Plan—the more household members the cheaper is the insurance per capita. Indeed, these mechanisms seem to cancel each other out, since there seems to be no heterogeneity by household size (Table C6).[43] The sliding scale of the KNCU Health Plan seems to be attractive for big households, since these are even more likely than small ones to take up insurance in the insurance intervention area (Table C6). Finally, again no test effect is found when *not* differentiating by test result (equation 3), for any of the outcomes (Table C7).

6.2. Test effect conditional on prior beliefs

Table 4 shows results corresponding to equation 2, when additionally controlling for self-reported HT at baseline (in the levels only). The test effects for the first three outcome variables are now more pronounced than in Table 2. The effect of normal BP on self-reported HT is now -12 pp (significant at the 0.1% level). Additionally, normal BP now has a negative effect on having had health provider consultations for HT in the past 12 months, as a priori expected, though it is not significant at the 10% level.

Table 4: Controlling for self-reported HT at baseline, corresponding to eq. 2

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
Normal BP	-0.116***	-0.037	-0.041	0.018
	(0.027)	(0.042)	(0.027)	(0.034)
High BP	0.187***	0.067	0.144***	-0.011
	(0.035)	(0.044)	(0.036)	(0.035)
Self-reported HT	-0.635***	-0.359***	-0.419***	0.018
	(0.031)	(0.038)	(0.039)	(0.027)
Constant	0.147***	0.120***	0.046*	0.059*
	(0.024)	(0.035)	(0.023)	(0.028)
Observations	3014	3006	3004	3020
P($\beta_{kn} = \beta_{kh}$)	<.001***	0.004	<.001***	0.237

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m= 12 months; $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Adding additional interaction terms "Normal BP × Self-reported HT" and "High BP × Self-reported HT" gives information on whether the test effect differs in case the result revealed unexpected instead of expected information. An interaction term of zero means that there is no significant difference between the two. Table 5 shows that all but one of the interaction terms are not significantly different from zero at the 10% level, contrary to Gong (2015), who finds that being surprised by the test result is the main driver of the behavioral response due to HIV testing. Only the parameter estimate for "Normal BP × Self-reported HT" is significantly different from zero, when the outcome is self-reported HT—in which case the interaction term is -21 pp (significant at the 5% level). Thus, someone who reported

[43] Running regression (4) of Table 3 at the household level, analogously to footnote 42, gives similar results.

to have HT at baseline but had normal BP is, as expected, much more likely to report not having HT at follow-up than someone unsurprised with normal BP.

From Table C8 of the Appendix it can be seen that, even when adding self-reported HT interaction terms to equation 4, there is no evidence of bias in the insurance impact estimates caused by the test.

Table 5: Heterogeneity by self-reported HT at baseline, corresponding to equation 2

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
Normal BP	-0.074**	-0.013	-0.042*	0.035
	(0.026)	(0.045)	(0.020)	(0.034)
Normal BP × Self-reported HT	-0.208*	-0.106	-0.018	-0.068
	(0.081)	(0.098)	(0.099)	(0.066)
High BP	0.185***	0.080	0.105***	0.017
	(0.038)	(0.050)	(0.031)	(0.034)
High BP × Self-reported HT	-0.012	-0.052	0.127	-0.098
	(0.081)	(0.093)	(0.099)	(0.064)
Self-reported HT	-0.551***	-0.297***	-0.463***	0.082+
	(0.065)	(0.075)	(0.077)	(0.046)
Constant	0.125***	0.104**	0.057**	0.042
	(0.024)	(0.037)	(0.019)	(0.027)
Observations	3014	3006	3004	3020

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

6.3. Robustness checks

Because of the age imbalance in the test treatment and control groups, the regressions are repeated, and reweighted such that in every age-group in both the insurance intervention and control areas the proportion of individuals who are assigned to the test treatment group is exactly 80%, as intended in the experimental design, see figure C4.[44] Regression results are robust to the reweighting. Those corresponding to Table 3 are shown in Table C9.

Results are also robust to sub-village level clustering of standard errors (Table C10). Furthermore, no heterogeneous test effect is found between the Hai and Moshi Rural treatment areas at a 5% level (Table C11).[45]

If the heterogeneity in behavioral response between normal and high BP is truly because of the baseline test result, this difference in response should be seen "on the margin" of the cut-off point in the definition of high BP—systolic BP $\geq$ 140 mmHg or diastolic BP $\geq$ 90 mmHg in two out of three measurements. This analysis will be done to alleviate any remaining concerns regarding the age imbalance in the sample. Note that the test control group cannot be included in these regressions. However note that this group was seen to behave similarly to the normal BP group for the two health care utilization outcomes, as well as for health

[44]The 12 individuals in the oldest age-group consist only of BP test individuals. This group thus cannot be reweighted and are dropped for the reweighted regressions.

[45]For brevity, robustness checks corresponding to other tables are not shown, but are available upon request.

insurance uptake—the regression coefficients for the normal BP group were consistently very close to zero.

The last three tables of Appendix C show the behavioral response on the margin, when only individuals are included with systolic (diastolic) BP within $140[1 \pm 0.x]$ ($90[1 \pm 0.x]$) for two out of three BP measurements, where x is taken to be 15, 10, and 5 (percent), respectively. This includes all high BP cases, and individuals who were only marginally away from being classified as such. These results indeed show a persistence of the previously seen behavioral response on the margin, confirming that the heterogeneity in health care seeking behavior between high BP and normal BP cases comes from the baseline BP measurement.

7. Conclusion

This paper provided strong evidence that individual behavior can be substantially influenced by the inclusion of rapid biomedical tests in surveys ("the test effect"). The test—in this case BP measurement—was embedded in an impact evaluation of a voluntary health insurance intervention in the Kilimanjaro region of Tanzania. This allowed to assess whether the test effect differed between the insurance intervention and control areas, something that would bias the insurance (intention to treat) impact estimates.

It was found that measured high BP significantly increased the likelihood to consult a health care provider for HT. The test effect did not differ between those surprised and those unsurprised with their BP measurement, as in Gong (2015). Unexpectedly, there was no test effect on health insurance enrollment, possibly because the insurance was offered at the household level. Finally, no evidence was found that the BP measurements biased health insurance impact estimates.

Medical clearance procedures mandate that results of rapid biomedical tests are revealed to the individuals who are tested during a survey. Considering the above evidence of behavioral change due to BP measurements, as well as that of earlier papers with malaria, HIV and water quality testing (e.g. Tarozzi et al., 2014; Thornton, 2008; Gong, 2015; Jalan and Somanathan, 2008; Hamoudi et al., 2012) it is recommended that the relatively easy and cheap method introduced in this paper, i.e. the random exclusion of a small percentage of interviewed individuals from participating in the test, should be routinely adopted in such surveys, to facilitate rigorous testing of whether these tests have biased the outcomes of interest.

References

SA Adams, CE Matthews, CB Ebbeling, CG Moore, JE Cunningham, J Fulton, and JR Hebert. The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology*, 161(4):389–398, 2005.

AIID and AIGHD. Impact Evaluation of HIF-supported Health Insurance Projects in Tanzania: Baseline Report KNCU Health Plan. `http://aiid.org/new/wp-content/uploads/2015/12/2007-03-Report-KNCU-Baseline-FINAL.pdf`, 2013.

JD Angrist and JS Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, 2009.

S Athey and G Imbens. Identification and inference in nonlinear difference-in-differences models. *The American journal of tropical medicine and hygiene*, 74(2):431–497, 2006.

WG Axinn, EA Jennings, and MP Couper. Response of sensitive behaviors to frequent measurement. *Social science research*, 49:1–15, 2015.

S Banoo, D Bell, P Bossuyt, A Herring, D Mabey, F Poole, PG Smith, N Sriram, C Wongsrichanalai, R Linke, et al. Evaluation of diagnostic tests for infectious diseases: general principles. *Nature Reviews Microbiology*, 8:S16–S28, 2008.

R Blundell and MC Dias. Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, 44(3):565–640, 2009.

RG Bridge, LG Reeder, D Kanouse, DR Kinder, V Tong Nagy, and CM Judd. Interviewing changes attitudes—sometimes. *Public Opinion Quarterly*, 41(1):56–64, 1977.

AC Cameron and PK Trivedi. *Microeconometrics: Methods and applications*. Cambridge university press, 2009.

TF Crossley, J de Bresser, L Delaney, and J Winter. Can survey participation alter household saving behavior? Technical report, IFS Working Papers, 2014.

J Das, J Hammer, and C Sánchez-Paramo. The impact of recall periods on reported morbidity and health seeking behavior. *Journal of Development Economics*, 98(1):76–88, 2012.

J Davis, AJ Pickering, K Rogers, S Mamuya, and AB Boehm. The effects of informational interventions on household water management, hygiene behaviors, stored drinking water quality, and hand contamination in peri-urban Tanzania. *The American journal of tropical medicine and hygiene*, 84(2):184–191, 2011.

A Delavande and H-P Kohler. The impact of HIV testing on subjective expectations and risky behavior in Malawi. *Demography*, 49(3):1011–1036, 2012. doi:10.1007/s13524-012-0119-7.

UM Dholakia. A critical review of question-behavior effect research. *Review of Marketing Research*, 7(7):145–197, 2010.

WH Dow, P Gertler, RF Schoeni, J Strauss, and D Thomas. *Health care prices, health and labor outcomes: Experimental evidence*. PSC Research Report No. 03-542, 2003.

E Duflo, R Glennerster, and M Kremer. Using randomization in development economics research: A toolkit. *Handbook of development economics*, 4:3895–3962, 2007.

JM Feldman and JG Lynch. Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of applied Psychology*, 73(3):421, 1988.

E Gong. HIV testing and risky sexual behaviour. *Economic Journal*, 125(582):32–60, 2015. doi:10.1111/ecoj.12125.

DR Gwatkin and A Ergo. Universal health coverage: friend or foe of health equity? *Lancet*, 377(9784):2160–2161, 2011.

DR Gwatkin, A Bhuiya, and CG Victora. Making health systems more equitable. *Lancet*, 364(9441):1273–1280, 2004.

A Hamoudi, M Jeuland, S Lombardo, S Patil, SK Pattanayak, and S Rai. The effect of water quality testing on household behavior: Evidence from an experiment in rural India. *The American journal of tropical medicine and hygiene*, 87(1):18–22, 2012.

ME Hendriks, FWNM Wit, TM Akande, B Kramer, GK Osagbemi, Z Tanović, E Gustafsson-Wright, LM Brewster, JMA Lange, and C Schultsz. Effect of health insurance and facility quality improvement on blood pressure in adults with hypertension in Nigeria: A population-based study. *JAMA internal medicine*, 174(4):555–563, 2014.

J Jalan and E Somanathan. The importance of being informed: Experimental evidence on demand for environmental quality. *Journal of Development Economics*, 87(1):14–28, 2008.

W Janssens, I de Beer, HM Coutinho, G van Rooy, J van der Gaag, and TF Rinke de Wit. Estimating HIV prevalence: A cautious note on household surveys in poor settings. *British Medical Journal*, 341:c6363, 2010. doi:10.1136/bmj.c6323.

W Janssens, J van der Gaag, TF Rinke de Wit, and Z Tanović. Refusal bias in the estimation of HIV prevalence. *Demography*, 51(3):1131–1157, 2014. doi:10.1007/s13524-014-0290-0.

KNCU. `http://www.kncutanzania.com/#!/page_company_profile`. Accessed: 2015-08-15.

CA Latkin and D Vlahov. Socially desirable response tendency as a correlate of accuracy of self-reported HIV serostatus for HIV seropositive injection drug users. *Addiction*, 93(8): 1191–1197, 1998. doi:10.1046/j.1360-0443.1998.93811917.x.

J Luoto, D Levine, and J Albert. Information and persuasion: achieving safe water behaviors in Kenya. RAND Working Paper Series No. WR-885, 2011.

NHIF. `http://www.nhif.or.tz/index.php/about-us/profile`. Accessed: 2015-11-24.

Oanda. `http://www.oanda.com/currency/historical-rates/`. Accessed: 2015-08-13.

OMRON. `http://www.omron-healthcare.com/en/products/bloodpressuremonitoring`. Accessed: 2015-11-25.

TG Pickering, GD James, C Boddie, GA Harshfield, S Blank, and JH Laragh. How common is white coat hypertension? *JAMA*, 259(2):225–228, 1988.

G Reniers and J Eaton. Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys. *AIDS*, 23(5):621629, 2009.

A Sen. Health: perception versus observation: self reported morbidity has severe limitations and can be extremely misleading. *British Medical Journal*, 324(7342):860–861, 2002.

SJ Sherman. On the self-erasing nature of errors of prediction. *Journal of personality and social psychology*, 39(2):211, 1980.

DE Sprott, ER Spangenberg, LG Block, GJ Fitzsimons, VG Morwitz, and P Williams. The question–behavior effect: What we know and where we go from here. *Social Influence*, 1 (2):128–137, 2006.

StataCorp. Stata statistical software: Release 11. College Station, TX: StataCorp LP, 2009.

Surveybe. `http://www.surveybe.com/index.php`.

A Tarozzi, A Mahajan, B Blackburn, D Kopf, L Krishnan, and J Yoong. Micro-loans, insecticide-treated bednets, and malaria: Evidence from a randomized controlled trial in Orissa, India. *American Economic Review*, 104(7):1909–41, 2014.

RL Thornton. The demand for, and impact of, learning HIV status. *American Economic Review*, 98(5):1829–1863, 2008. doi:10.1257/aer.98.5.1829.

RL Thornton. HIV testing, subjective beliefs and economic behavior. *Journal of Development Economics*, 99(2):300–313, 2012. doi:10.1016/j.jdeveco.2012.03.001.

CG Victora, K Hanson, J Bryce, and JP Vaughan. Achieving universal coverage with health interventions. *Lancet*, 364(9444):1541–1548, 2004.

JM Wooldridge. *Econometric analysis of cross section and panel data*. The MIT press, 2002.

World Bank. `http://data.worldbank.org/indicator/PA.NUS.PPPC.RF`. Accessed: 2015-08-15.

AP Zwane, J Zinman, E Van Dusen, W Pariente, C Null, E Miguel, M Kremer, DS Karlan, R Hornbeck, X Giné, E Duflo, F Devoto, B Crepon, and A Banjeree. Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences*, 108(5):1821–1826, 2011.

Appendix A Behavioral response to household survey testing in the literature

A.1 HIV testing

Thornton (2008) conducted an experiment in rural Malawi where people were randomly assigned monetary incentives to learn their HIV status after having been tested during a household survey. She finds that sexually active HIV positive individuals who learn their status are three times more likely to buy condoms two months later, compared to those who did not learn their status. She finds no test effect on condom purchases for HIV negative individuals. Based on the same experiment, Thornton (2012) finds no effect of learning one's HIV status on economic behavior two years later. Delavande and Kohler (2012), also using data from this experiment, observe surprisingly that becoming aware of an HIV negative status increases one's subjective beliefs of being HIV positive after two years. Those who learned they were HIV positive self-reported less risky sexual behavior—more condom use and less sexual partners—than HIV positive individuals who did not learn their result.

Using data from experiments in urban Kenya and Tanzania where HIV tests were randomly assigned during a survey, Gong (2015) finds that individuals *surprised* by their HIV positive (negative) result had higher (lower) likelihood to contract gonorrhea or chlamydia—sexually transmitted infections (STIs), and proxies of risky sexual behavior—six months after the experiment, compared to untested individuals with similar prior beliefs about their HIV infection probability. However, no test effect on STI infection was observed for those unsurprised by their test result.

A.2 Water quality testing

Jalan and Somanathan (2008) conducted a household survey in urban India, where households' own drinking water was tested for the presence of fecal bacteria, and a random sample learned their test result. Households initially not purifying their water that received bad news—denoting presence of fecal bacteria in their drinking water—were more likely to start purifying their water within two months of learning the result, compared to those not learning their test result. Initially purifying households that received good news did not reduce purification.

Davis et al. (2011) use data from a household survey experiment in peri-urban Tanzania where all households received general information on water quality and treatment, and randomly chosen households were notified of their own stored drinking water and/or hand-rinse water quality. Households that received the test results—all of which were contaminated—were more likely to self-report improved water management and hygiene behaviors, but less likely to improve their actual drinking water or hand-rinse water quality two months later, compared to the control group—which received general information only.

Luoto et al. (2011) tested both source water and households' own water quality in rural Kenya in the scope of a household survey. All households received water purification kits, and a random sub-sample learned that their source water was contaminated. Furthermore, a random sub-sample of the latter additionally received the test result of their household's own stored water—most times (87%) contaminated. The authors find that information provision of source water quality significantly increases water purification efforts two months later. These do not increase further once the household's own water quality is revealed.

In a household survey in rural India a random sample of households learned the quality of their own drinking water and received information on water management practices, while the control group received neither test nor information. Using data from this experiment Hamoudi et al. (2012) find, in a situation where almost 90% of tested households had contaminated drinking water, that water quality testing increased demand for commercial water sources, but had no effect on time-intensive water management practices.

Appendix B Sampling

In November 2012 to January 2013 a census was conducted of all households belonging to active KNCU members of the insurance intervention and control areas in districts Hai, Moshi Rural, and Rombo of the Kilimanjaro region. A random sample of 1,500 households was then selected from the census, stratified by geographic area and insurance intervention.[46] Namely 500 households were drawn from the Rombo district and 1,000 from the Hai and Moshi Rural districts (the study population in Moshi Rural district is near the border with Hai district), such that half of the sample in each of these areas was drawn from the insurance intervention area, and the other half from the insurance control area. Because of logistical reasons an additional stratification was made at the smallest administrative unit, the sub-village, such that from each sub-village, in each stratum, approximately the same number of households was randomly drawn. Thus, sampling probability weights are necessary in the analyses.

Additionally, per sub-village, 40% extra households were randomly drawn, to serve as replacement in case a household in the original sample could not be interviewed. Ultimately 84 households (5.6%) of the original sample were replaced (equally distributed between the insurance treatment and control group), mostly due to households' unavailability for the interview (39%), followed by: household moved (17%), household listed twice (17%), household refused to be interviewed (16%), household unknown (6%), household no longer exists (6%).

[46] The census data were furthermore used by the KNCU Health Plan as administrative base for the insurance.

Appendix C Additional tables and figures

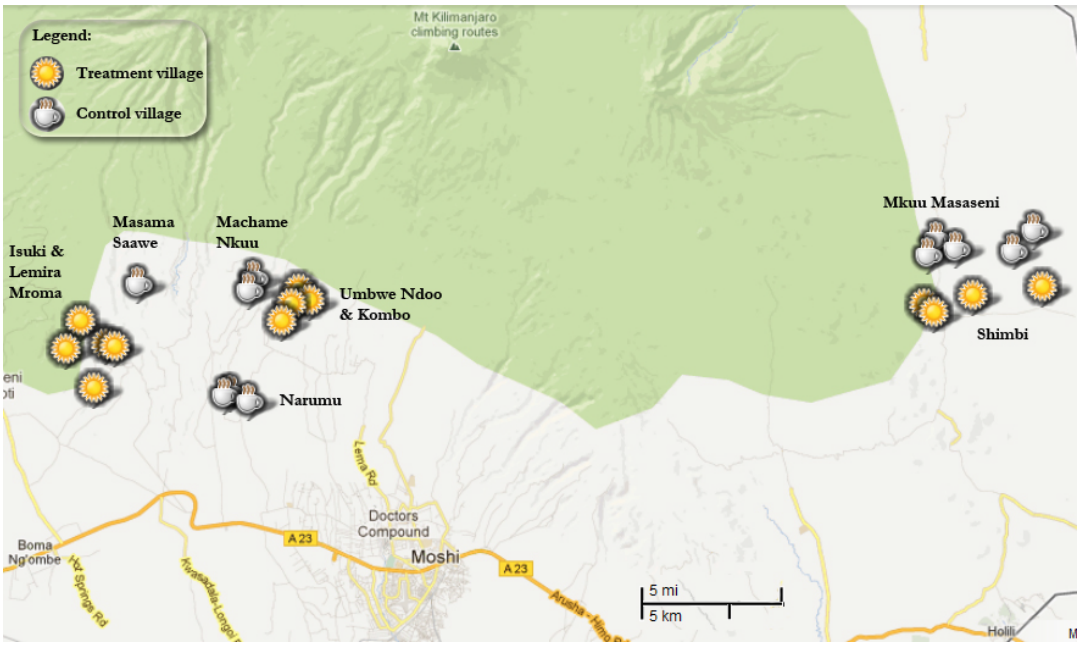

Figure C1: Insurance intervention and control KNCU primary societies (PSs) and their villages. *Hai* district is located on the far west side of the map (containing PSs Isuki, Lemira Mroma, Masama Saawe, Machame Nkuu, and Narumu). *Moshi Rural* is the district east of Hai (containing PSs Umbwe Ndoo and Kombo), while *Rombo* is the most eastern district (containing PSs Shimbi and Mkuu Masaseni). For completion, Moshi Urban district is located to the south of Moshi Rural, and is not included in the research. The (control) treatment villages in this figure denote the areas where—at the time of the baseline survey—the health insurance intervention, i.e. the KNCU Health Plan, was *planned* (not) to be offered. Because it proved not possible to introduce the KNCU Health Plan in the Rombo district, the two PSs in Rombo were excluded from the follow-up survey and subsequent analyses. Note that the research population resides on the slopes of mount Kilimanjaro, increasing in altitude towards the north-northwest of the map. Substantial travel time is needed between treatment and control group villages, sometimes even between villages of the same primary society. Lemira Lutheran dispensary is located within the cluster of Isuki and Lemira Mroma villages (north side). Umbwe Parish dispensary is located in the Umbwe Ndoo and Kombo village cluster (southeast side). PSs Masama Saawe, Machame Nkuu, and Narumu have a popular dispensary within their village cluster as well. Source: GPS coordinates collected by EDI Ltd. in November 2012. Map adapted from Google maps.

Table C1: KNCU Health Plan details per primary society (insurance intervention group)

Primary society	District	Primary health facility	Start date	End date
Isuki	Hai	Lemira Lutheran dispensary^{+}	1-Jul-2013	30-Jun-2014*
Lemira Mroma	Hai	Lemira Lutheran dispensary^{+}	1-Aug-2013	31-Jul-2014*
Umbwe Ndoo	Moshi Rural	Umbwe Parish dispensary	1-Oct-2013	31-Dec-2014**
Kombo	Moshi Rural	Umbwe Parish dispensary	1-Oct-2013	31-Dec-2014**
Shimbi	Rombo	n/a	n/a	n/a

Note: $^{+}$Lemira Lutheran dispensary was upgraded in the scope of the KNCU Health Plan in May–August 2013. *In Hai district the KNCU Health Plan could be renewed for another year after the shown end date. **In Moshi Rural the first insurance period was extended until the end of 2014 (three additional months for free). As of 1 January 2015 the KNCU Health Plan was terminated in this district. Instead, all households in Moshi Rural could then enroll in the new *i*CHF insurance (see Section 2).

Table C2: Attrition: baseline means of sample vs. individuals lost to follow-up

	Sample Mean (N=1536)	Lost to follow-up Mean (N=623)	ΔMean p-value
Main			
BP test	0.81	0.80	0.841
High BP[a]	0.31	0.21	<.001***
Insurance program area	0.50	0.41	<.001***
Self-reported HT	0.23	0.14	<.001***
BP check - past 12 months	0.35	0.26	<.001***
Consult for HT - past 12 months	0.17	0.12	<.001***
Any health insurance	0.15	0.14	0.575
Socio-economic characteristics			
Age (years)	55.4	43.1	<.001***
Female	0.60	0.48	<.001***
Married[b]	0.69	0.42	<.001***
Worked[c] - past 12 months	0.20	0.25	0.008**
Religion: Christian	0.96	0.97	0.475
Mother tongue: Chagga	0.99	0.99	0.160
Educ[d]: None	0.10	0.10	0.997
Educ: Less than primary school	0.31	0.18	<.001***
Educ: Primary school	0.53	0.57	0.148
Educ: More than primary school	0.06	0.16	<.001***
Self-reported illness/ injury			
Chronic illness	0.42	0.28	<.001***
Acute illness / injury - past 12 months	0.50	0.42	<.001***
Hospitalization - past 12 months	0.07	0.07	0.931
Household characteristics			
Annual consumption[e] - PC (TZS / 1,000)	862	862	0.998
Financial health shock - past 12 months	0.37	0.41	0.087^{+}
Household size	4.34	4.86	<.001***
#Young children in HH (age < 5)	0.19	0.20	0.758
#Elderly in HH (age ≥ 60)	0.73	0.61	<.001***
#Reproductive age women (15–45) in HH	0.50	0.62	<.001***

Note: The table shows statistics for all adults who gave consent for the biomedical part of the baseline survey (questions and measurements). Means weighted and p-values clustered at the household level, in accordance with the sampling method HH=household; PC= per capita; [a]43 individuals with missing test result excluded (Sample: 25; Lost to follow-up: 18); [b]Includes mono- and polygamous marriage; [c]Apart from household chores and family farming; [d]Highest completed education level; [e]Two outliers excluded; $^{+}$ p<.10, * p<.05, ** p<.01, *** p<.001.

Table C3: Baseline means of individuals lost to follow-up, by test assignment

	BP Test Mean (N=495)	No BP test Mean (N=128)	ΔMean p-value
Main			
Insurance program area	0.41	0.44	0.613
Self-reported HT	0.14	0.15	0.813
BP check - past 12 months	0.27	0.21	0.214
Consult for HT - past 12 months	0.11	0.13	0.738
Any health insurance	0.13	0.17	0.392
Socio-economic characteristics			
Age (years)	43.1	42.8	0.892
Female	0.50	0.40	0.108
Married[a]	0.42	0.40	0.667
Worked[b] - past 12 months	0.24	0.28	0.439
Religion: Christian	0.97	0.97	0.989
Mother tongue: Chagga	0.99	0.99	0.686
Educ[c]: None	0.10	0.09	0.710
Educ: Less than primary school	0.17	0.21	0.399
Educ: Primary school	0.58	0.49	0.090^{+}
Educ: More than primary school	0.14	0.21	0.150
Self-reported illness/ injury			
Chronic illness	0.27	0.31	0.485
Acute illness / injury - past 12 months	0.42	0.43	0.894
Hospitalization - past 12 months	0.07	0.09	0.497
Household characteristics			
Annual consumption[d] - PC (TZS / 1,000)	857	883	0.629
Financial health shock - past 12 months	0.42	0.38	0.578
Household size	4.90	4.70	0.385
#Young children in HH (age $<$ 5)	0.21	0.15	0.299
#Elderly in HH (age $\geq$ 60)	0.60	0.64	0.645
#Reproductive age women (15–45) in HH	0.64	0.55	0.349

Note: The table shows statistics for all adults who gave consent for the biomedical part of the baseline survey (questions and measurements), but who were lost to follow-up. Means are weighted and p-values clustered at the household level, in accordance with the sampling method. HH= household; PC=per capita; [a]Includes mono- and polygamous marriage; [b]Apart from household chores and family farming; [c]Highest completed education level; [d]One outlier excluded; $^{+}$ p<.10, * p<.05, ** p<.01, *** p<.001.

Table C4: Baseline characteristics, by insurance intervention/control area

	Ins. area Mean (N=797)	Not Ins. area Mean (N=739)	ΔMean p-value
Main			
BP test	0.79	0.82	0.188
High BP[a]	0.25	0.37	<.001***
Self-reported HT	0.22	0.25	0.154
BP check - past 12 months	0.33	0.37	0.060^{+}
Consult for HT - past 12 months	0.17	0.17	0.685
Any health insurance	0.15	0.14	0.743
Socio-economic characteristics			
Age (years)	54.9	55.8	0.307
Female	0.60	0.61	0.712
Married[b]	0.70	0.69	0.666
Worked[c] - past 12 months	0.20	0.20	0.808
Religion: Christian	0.94	0.98	0.003**
Mother tongue: Chagga	1.00	0.97	0.002**
Educ[d]: None	0.11	0.09	0.176
Educ: Less than primary school	0.31	0.32	0.711
Educ: Primary school	0.54	0.52	0.438
Educ: More than primary school	0.04	0.07	0.003**
Self-reported illness/ injury			
Chronic illness	0.41	0.44	0.252
Acute illness / injury - past 12 months	0.49	0.51	0.463
Hospitalization - past 12 months	0.08	0.07	0.337
Household characteristics			
Annual consumption[e] - PC (TZS / 1,000)	851	873	0.382
Financial health shock - past 12 months	0.39	0.36	0.432
Household size	4.27	4.40	0.340
#Young children in HH (age <5)	0.18	0.16	0.486
#Elderly in HH (age ≥ 60)	0.72	0.72	0.893
#Reproductive age women (15–45) in HH	0.50	0.49	0.841

Note: The table shows statistics for all adults who gave consent for the biomedical part of both surveys (questions and measurements). Means are weighted and p-values clustered at the household level, in accordance with the sampling method. HH=household; PC=per capita; [a]25 individuals with missing test result excluded (Ins. Treat: 8; Ins. Control: 17); [b]Includes mono- and polygamous marriage; [c]Apart from household chores and family farming; [d]Highest completed education level; [e]One outlier excluded; Ins.=Insurance intervention; $^{+}$ p<.10, * p<.05, ** p<.01, *** p<.001.

Table C5: Results corresponding to equation 1

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
BP test	0.022	0.014	0.044	0.009
	(0.032)	(0.039)	(0.030)	(0.032)
Constant	-0.018	0.029	-0.064*	0.063*
	(0.028)	(0.034)	(0.027)	(0.028)
Observations	3064	3056	3056	3072

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

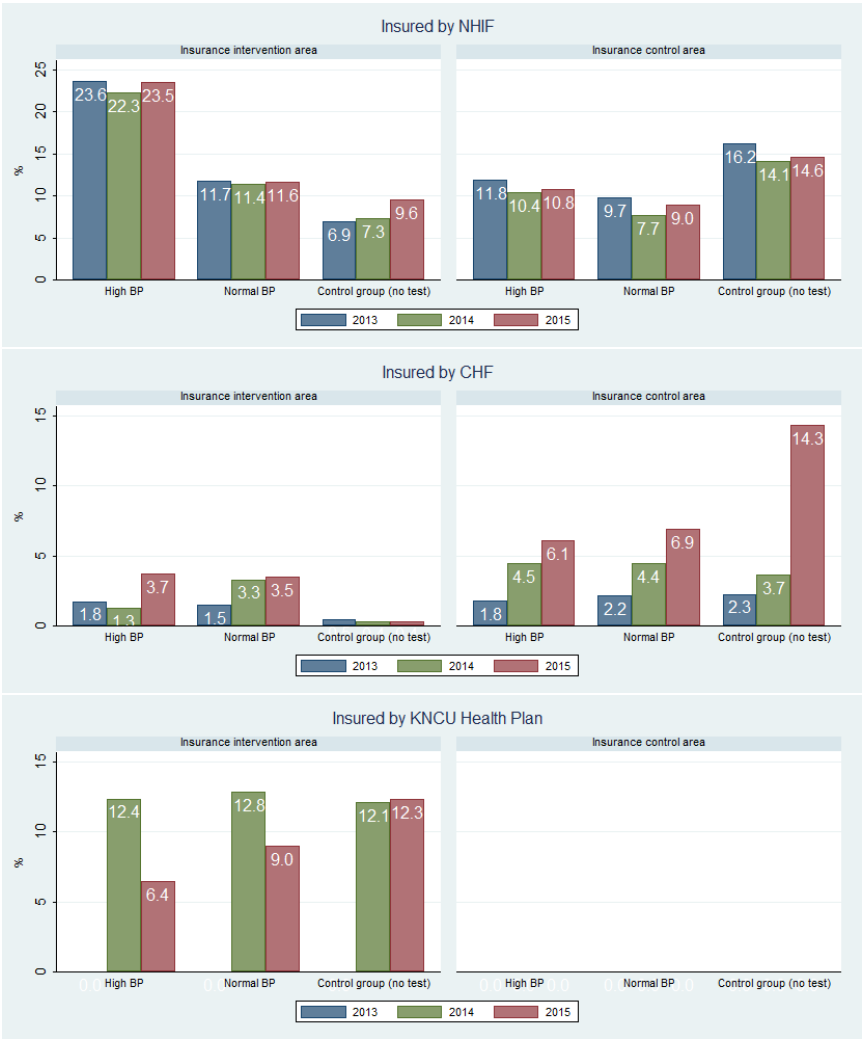

Figure C2: Detailed health insurance enrollment, by test result and insurance area (N=1511). No one in the insurance control area took up the KNCU Health Plan, as envisaged in the research design. Observations in the insurance intervention [control] area: Total: 789 [722]; High BP: 209 [258]; Normal BP: 425 [326]; Control group (no test): 155 [138].

Figure C3: Average outcome by baseline age-group, test result, and year (N=1511). The baseline and follow-up lines are roughly parallel, indicating no heterogeneity in the test effect by age. Number of observations per test result: High BP: 467; Normal BP: 751; Control group (no test): 293. High BP age-cohort 18–39 has the smallest number of observations (16), followed by the same age-cohort of the control (no test) group (43).

Table C6: Heterogeneity by household size at baseline

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
Normal BP	-0.041	-0.009	0.012	-0.005
	(0.067)	(0.073)	(0.064)	(0.060)
Normal BP × Big HH	0.027	-0.002	-0.027	0.038
	(0.096)	(0.112)	(0.082)	(0.090)
High BP	0.192*	0.082	0.169*	0.014
	(0.074)	(0.080)	(0.074)	(0.060)
High BP × Big HH	-0.054	-0.043	-0.114	0.027
	(0.112)	(0.116)	(0.098)	(0.091)
Normal BP × Ins.area	-0.013	-0.052	-0.012	0.085
	(0.088)	(0.106)	(0.083)	(0.082)
Normal BP × Ins.area × Big HH	-0.071	0.185	0.047	-0.199
	(0.130)	(0.172)	(0.123)	(0.144)
High BP × Ins.area	-0.013	-0.120	-0.021	-0.020
	(0.108)	(0.115)	(0.106)	(0.087)
High BP × Ins.area × Big HH	-0.107	0.174	0.040	-0.069
	(0.157)	(0.182)	(0.155)	(0.149)
Ins.area	0.055	0.046	0.070	0.052
	(0.075)	(0.085)	(0.073)	(0.064)
Ins.area × Big HH	0.000	-0.141	-0.100	0.213+
	(0.111)	(0.140)	(0.109)	(0.122)
Big HH	0.004	-0.007	0.039	-0.087
	(0.082)	(0.087)	(0.070)	(0.078)
Constant	-0.050	0.031	-0.099+	0.030
	(0.058)	(0.060)	(0.057)	(0.051)
Observations	3014	3008	3006	3022

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; Ins.area=Insurance intervention area; HH=household; Big HH=HH with more than 4 members; $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table C7: Results corresponding to equation 3

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
BP test	0.064	0.025	0.056	0.018
	(0.047)	(0.051)	(0.043)	(0.044)
BP test × Ins.area	-0.079	-0.023	-0.021	-0.002
	(0.064)	(0.077)	(0.061)	(0.064)
Ins.area	0.055	0.001	0.034	0.128*
	(0.056)	(0.067)	(0.054)	(0.055)
Constant	-0.048	0.028	-0.082*	-0.006
	(0.041)	(0.044)	(0.038)	(0.039)
Observations	3064	3056	3056	3072

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; Ins.area=Insurance intervention area; $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$.

Table C8: Heterogeneity by self-reported HT at baseline

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
Normal BP	-0.081*	-0.005	-0.065*	0.009
	(0.040)	(0.066)	(0.029)	(0.052)
Normal BP × Self-reported HT	-0.155	-0.138	0.019	0.007
	(0.116)	(0.120)	(0.130)	(0.088)
High BP	0.177***	0.107	0.067+	0.036
	(0.053)	(0.069)	(0.040)	(0.050)
High BP × Self-reported HT	-0.011	-0.116	0.183	-0.040
	(0.113)	(0.116)	(0.125)	(0.086)
Normal BP × Ins.area	0.013	-0.013	0.041	0.052
	(0.053)	(0.090)	(0.040)	(0.068)
Normal BP × Ins.area × Self-reported HT	-0.108	0.078	-0.070	-0.171
	(0.163)	(0.199)	(0.199)	(0.132)
High BP × Ins.area	0.012	-0.048	0.093	-0.004
	(0.078)	(0.102)	(0.064)	(0.074)
High BP × Ins.area × Self-reported HT	0.008	0.122	-0.134	-0.158
	(0.162)	(0.188)	(0.202)	(0.131)
Ins.area	-0.019	0.045	-0.011	0.094+
	(0.049)	(0.074)	(0.037)	(0.055)
Ins.area × Self-reported HT	0.089	-0.244	0.018	0.161+
	(0.132)	(0.153)	(0.157)	(0.094)
Self-reported HT	-0.594***	-0.180*	-0.472***	0.019
	(0.092)	(0.086)	(0.095)	(0.062)
Constant	0.136***	0.079	0.063*	-0.012
	(0.037)	(0.054)	(0.027)	(0.045)
Observations	3014	3006	3004	3020

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; Ins.area= Insurance intervention area; + $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

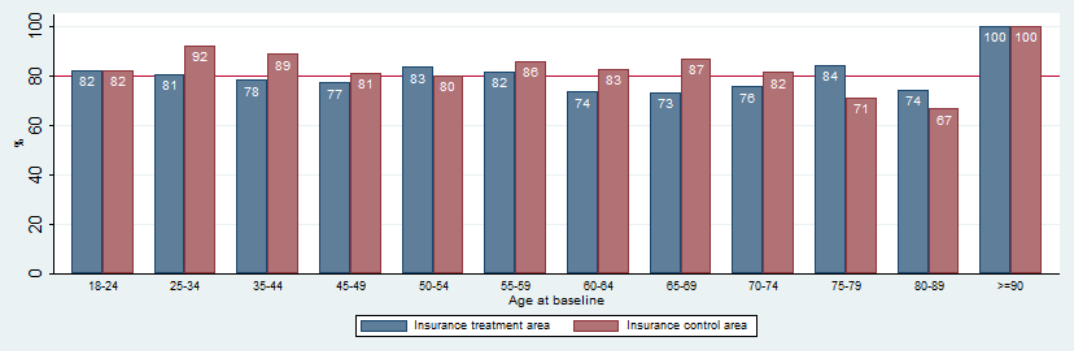

Figure C4: Age-group reweighting, based on baseline age. Bins of five years were grouped together (18–24, 25–34, 35–44, 80–89) if they had less than five test control group individuals, to avoid assigning extremely high weights to single individuals. The oldest individual in the test control group is 89 years old, which is why the ≥ 90 group is 100% treated. The 12 individuals in this age group are thus excluded from the reweighted regressions.

Table C9: Reweighted by agegroup

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
Normal BP	-0.058	-0.025	-0.030	0.012
	(0.046)	(0.058)	(0.038)	(0.047)
High BP	0.145**	0.042	0.090+	0.028
	(0.054)	(0.060)	(0.047)	(0.048)
Normal BP × Ins.area	-0.010	0.038	0.026	0.017
	(0.061)	(0.082)	(0.056)	(0.067)
High BP × Ins.area	-0.024	-0.027	0.022	-0.036
	(0.076)	(0.088)	(0.074)	(0.070)
Ins.area	0.017	-0.031	0.003	0.128*
	(0.051)	(0.065)	(0.048)	(0.055)
Constant	-0.019	0.043	-0.051	-0.010
	(0.038)	(0.047)	(0.032)	(0.041)
Observations	2988	2984	2982	2996
P($\tilde{\beta}_{kn} = \tilde{\beta}_{kh}$)	<.001***	0.181	0.003**	0.509
P($\tilde{\eta}_{kn} = \tilde{\eta}_{kh}$)	0.832	0.386	0.955	0.282

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; Ins.area= Insurance intervention area; $^{+}p < .10$, $^{*}\ p < .05$, $^{**}\ p < .01$, $^{***}\ p < .001$.

Table C10: Cluster by subvillage

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
Normal BP	-0.029	-0.010	0.001	0.008
	(0.046)	(0.058)	(0.036)	(0.022)
High BP	0.169*	0.064	0.121**	0.024
	(0.064)	(0.045)	(0.041)	(0.027)
Normal BP × Ins. area	-0.044	0.013	-0.000	0.016
	(0.068)	(0.087)	(0.056)	(0.045)
High BP × Ins. area	-0.053	-0.061	-0.006	-0.037
	(0.076)	(0.062)	(0.061)	(0.064)
Ins. area	0.055	0.001	0.034	0.128***
	(0.056)	(0.073)	(0.048)	(0.034)
Constant	-0.048	0.028	-0.082**	-0.006
	(0.041)	(0.046)	(0.030)	(0.017)
Observations	3014	3008	3006	3022
P($\tilde{\beta}_{kn} = \tilde{\beta}_{kh}$)	<.001***	0.238	0.008**	0.482
P($\tilde{\eta}_{kn} = \tilde{\eta}_{kh}$)	0.893	0.350	0.910	0.303

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; Ins.area= Insurance intervention area; $^{+}p < .10$, sym$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table C11: Heterogeneity by insurance intervention area

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
Normal BP	-0.029	-0.010	0.001	0.008
	(0.048)	(0.056)	(0.043)	(0.046)
High BP	0.169**	0.064	0.121*	0.024
	(0.056)	(0.058)	(0.051)	(0.046)
Normal BP × Ins. area	-0.056	0.031	0.019	0.006
	(0.072)	(0.094)	(0.069)	(0.074)
Normal BP × Moshi Rural	0.028	-0.105	-0.088	0.018
	(0.085)	(0.122)	(0.077)	(0.107)
High BP × Ins. area	-0.023	-0.028	0.040	-0.014
	(0.092)	(0.103)	(0.091)	(0.081)
High BP × Moshi Rural	-0.107	-0.155	-0.174+	-0.094
	(0.105)	(0.128)	(0.103)	(0.112)
Ins. area	0.046	-0.036	0.015	0.113+
	(0.061)	(0.075)	(0.060)	(0.059)
Moshi Rural	0.040	0.172+	0.088	0.066
	(0.075)	(0.103)	(0.069)	(0.090)
Constant	-0.048	0.028	-0.082*	-0.006
	(0.042)	(0.044)	(0.038)	(0.040)
Observations	3014	3008	3006	3022

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; Ins.area= Insurance intervention area; $^{+}$ $p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table C12: 15% margin

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
High BP	0.163**	0.044	0.089+	0.020
	(0.050)	(0.062)	(0.046)	(0.029)
High BP × Ins. area	0.022	-0.010	0.037	-0.050
	(0.070)	(0.088)	(0.068)	(0.057)
Ins. area	-0.004	-0.044	0.022	0.127**
	(0.041)	(0.065)	(0.036)	(0.044)
Constant	-0.049	0.058	-0.064*	0.005
	(0.033)	(0.047)	(0.028)	(0.026)
Observations	1788	1784	1782	1790

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; Ins.area= Insurance intervention area; $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table C13: 10% margin

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
High BP	0.173**	0.046	0.106*	0.009
	(0.055)	(0.065)	(0.052)	(0.033)
High BP × Ins. area	-0.002	-0.050	-0.014	-0.039
	(0.076)	(0.097)	(0.073)	(0.065)
Ins. area	0.041	0.005	0.069	0.102+
	(0.049)	(0.074)	(0.046)	(0.054)
Constant	-0.080*	0.037	-0.087*	0.020
	(0.039)	(0.051)	(0.037)	(0.033)
Observations	1442	1438	1436	1444

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted). 12m=12 months; Ins.area= Insurance intervention area; $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table C14: 5% margin

	(1) Self-reported HT	(2) BP check: 12m	(3) Consult for HT: 12m	(4) Insured
High BP	0.099	0.125	0.040	-0.023
	(0.061)	(0.099)	(0.056)	(0.046)
High BP × Ins. area	0.063	-0.121	0.021	0.009
	(0.089)	(0.143)	(0.086)	(0.070)
Ins. area	-0.028	0.030	0.023	0.064
	(0.063)	(0.120)	(0.058)	(0.069)
Constant	0.009	-0.031	-0.018	0.035
	(0.042)	(0.085)	(0.038)	(0.048)
Observations	876	870	870	876

Note: Fixed effects regression on balanced two-period panel; weighted and standard errors clustered at household level, according to the sampling frame. Standard errors in parentheses. All independent variables are interacted with time dummy (notation "× Time" omitted); 12m=12 months; Ins.area= Insurance intervention area; $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.